American Journal of Medical Genetics (Neuropsychiatric Genetics) 67:1–8 (1996)

Reliable Computer-Assisted Classification of the EEG: EEG Variants in Index Cases and Their First Degree Relatives

R.M. Dünki, G.B. Schmid, P. Scheidegger, H.H. Stassen, G. Bomben, and P. Propping

Psychiatric University Hospital, Research Department, Zürich, Switzerland (R.M.D., G.B.S., P.S., H.H.S., G.B.) and University of Bonn, Department of Human Genetics, Bonn, Germany (P.P.)

A method which optimizes on global properties of sample recordings is proposed for the definition of and the discrimination between electroencephalogram (EEG) classes. The sample was drawn from students at the University of Heidelberg from 1974 to 1978 and consists of 15 healthy index cases clinically ascertained as belonging to the low voltage EEG group. In addition, the three clinically defined groups: diffuse β (18 index cases), borderline α (12 index cases) and monomorphous α (18 index cases) have been included in the study, as well as the first degree relatives of the index cases, thus providing a clinical classification into four groups. The proposed method provides an automatic and reliable classification algorithm using discriminant and cluster analysis. The relation between such an automatized classification and clinical classification schemes is investigated. In particular, the inheritance of the low voltage EEG, the question on sex differences and the question of a simple Mendelian mechanism had been examined.

The method of random splittings had been applied for discriminant and cluster analysis. Our findings can be summarized as follows: (1) except for the monomorphous α EEG group, the clinical classification shows rather marginal separation (discriminating performance 60% to 75%), while a new and more reliable grouping scheme improves the discriminating performance up to 87% to 91%. The latter scheme leads to the concept of personal channel pattern (PCP) and was compared to the clinical classification scheme by means of contingency tables; (2) only a weak correlation between the clinically and PCP-based groups could be found (Cramér Index: 0.27). Accordingly, we continued to investigate the extent to which the proposed EEG classification scheme can nevertheless explain the genetic mechanisms apparently involved in the low voltage EEG. We thus considered the role of sex differences manifest in our proposed new grouping scheme; (3) males occurred more frequently in the new group 3 and females more frequently in the new group 1. In this regard, a much better correlation of the new groups between mothers and children than between fathers and children was observed; and (4) with help of our new PCP scheme, we have been able to reproduce a simple two gene Mendelian scheme to explain inheritance of the clinical low voltage EEG group. In this PCP-based scheme, the low voltage property does not occur when dominance of a certain gene (called gene A) is absent.

KEY WORDS: discriminant analysis, cluster analysis, personal channel pattern, heritability, Mendelian interpretation

INTRODUCTION

The human electroencephalogram (EEG) is clinically classified into different types. These types possess different signal characteristics, e.g., frequency or amplitude domain. One of these types is characterised by the so-called low voltage EEG (Fig. 1) which interestingly seems to follow a Mendelian, namely, autosomal-dominant mode of inheritance [Vogel, 1986]; the existence of this variant in a given individual implies that at least one parent also carries a low voltage EEG. We emphasize that this genetic finding is intrinsic to the clinical classification scheme.

A central problem inherent in genetic studies of the EEG is the reliability of classifying EEG variants. Reliability means independence of the results from the clinician's personal judgement and (in a statistical sense)

Received for publication July 8, 1994; revision received June 29, 1995.

Address reprint requests to Dr. R.M. Dünki, Computer Assisted Physics Group, University Zürich Irchel, Winterthurestrasse 190, CH-8057 Zürich, Switzerland.

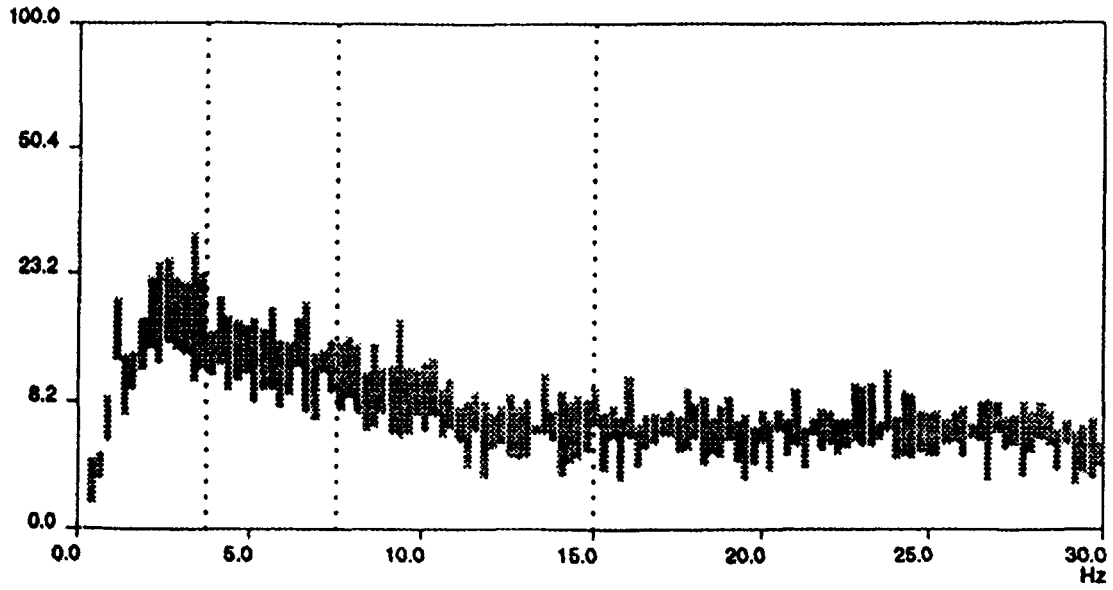

Fig.1. Spectral pattern of a clinically classified low voltage EEG. Dotted lines represent borderlines of the four frequency bands used for analysis.

from the sampling of probands. This problem has not yet been definitively solved [Krueger et al., 1981].

Accordingly, one must find a reliable classification scheme specific to EEG variants and incorporate this scheme into an appropriate algorithm to distinguish between classes. The search for such a scheme comprised the first step of our work. Having found an appropriate algorithm in the second step, the following questions remained: (a) is the distinction between the EEG variants sufficiently compatible with the clinical classification scheme or must a new scheme be introduced? (b) if the new scheme does not correlate well with the old scheme, do we still find evidence for a genetic mechanism within the new one? (c) given such evidence, are there specific preferences, e.g., specific parent-offspring correlations or male/female differences? and (d) is there evidence for an interpretation in terms of Mendelian genetics? In particular, is one able to confirm the genetic evidence of the old scheme within the scope of the new one by means of a model in which only one or a few genes with dominant and recessive alleles can be considered?

The remainder of this paper is organized as follows. First, we examine the power of our method for classifying EEG variants and compare the result with the clinical typing of EEG classes. Secondly, we explore the evidence for the genetic control of the EEG variation including the possibility of sex differences. Thirdly, we provide a confirmation of the inheritance of the EEG variants, in agreement with the usual Mendelian arguments, but independent of the clinical classification scheme.

MATERIALS AND METHODS

Sample Description

This study is based on EEG recordings of students and their first degree relatives (siblings and parents) sampled in Heidelberg from 1974 to 1978, and the data have already been used for a family study. Four groups of carriers of EEG variants were delineated: 15 index cases with low voltage EEG (group 1, ages between 23 and 37 years), 18 cases with diffuse β-waves (group 2), 12 cases with borderline α (group 3) and 18 cases with monomorphous α-waves (group 4). The classification schemes (cf. below) have been derived solely from these index cases. For the genetic study, the 37 first degree relatives of the index group 1 have also been classified with these schemes and used in the analysis.

Data have been recorded analog under the condition of quiet wakefulness on the channels: F3, F4, C3, C4, P3, P4, O1, and O2 according to the international 10–20 scheme [Cooper et al., 1969] and were labelled internally in this order as channel 1, channel 2, . . . , to channel 7, channel 0. The right ear was used as reference point for measurements on the right hemisphere and the left ear for measurements on the left hemisphere. Data have been digitized afterwards at a sampling rate of 256 Hz with 10-bit resolution.

Procedure

The recorded EEG was split into successive segments of 20-second lengths. These EEG segments were checked with an internal artefact code [Stassen et al., 1987]. We did not use segments with an artefact code higher than 69 (a maximum of 8-sec time series were "blanked out" during artifact processing). Channels with less than three usable segments had to be excluded from analysis because no reliable averages of the parameters could be estimated. Circa 82% of all the data remained within the sample after this procedure. On a per person basis this means that, on average, 6.6 channels were available per person.

Analysis of EEG can be done either in the time domain or in the frequency domain. The latter provides

excellent results for investigations within the genetic score [Stassen et al., 1987, 1988]. Analysis in the frequency domain has been developed, standardized and automatized to a high degree. Therefore our analysis was performed in the frequency domain.

In addition to the original parameters, relative variables such as relative peak frequency and relative peak amplitude had been included into the feature vectors. A relative variable is defined by $X_{rel}(I) := X(I)/X(1)$, where X(I) is either the peak frequency or the peak amplitude within frequency band I. The introduction of these relative variables has been motivated by picturing the EEG to result from an oscillating system [i.e., Wright, 1990]. For an oscillating system, the peak parameters are not independent of each other, so that they display a smaller deviation around their mean than do truly independent variables.

Spectral analysis was carried out following a tonal scheme [Stassen, 1991]. Unless stated otherwise, the four bands were set to 1: 0–3.75 Hz; 2: 3.75–7.50 Hz; 3: 7.50–15.00 Hz; 4: 15.00–32.00 Hz for our analysis. From each band, the six parameters absolute power, relative power, centroid, symmetry, peak amplitude and peak frequency were selected [Stassen and Bomben, unpublished]. Relative peak frequency and relative peak amplitude were also selected for the upper three frequency bands. In addition, we used the total power over the whole spectrum. This provides us with a total of 31 variables which initially made up a 31-dimensional feature vector.

Computational Procedures

Computer analysis was performed with the MASTER.EEG [Stassen, 1992], LISREL [Jöreskog and Sörbom, 1989] and SAS statistical software packages. Discriminant and cluster procedures followed the method of random splittings and are described in more detail below.

The search for evidence of a genetic influence was performed by comparing the low voltage EEG index cases and their relatives. Haldane-Dawson and Yates small sample corrections were applied, if necessary. Some of the tests were based on our classification scheme, while in other tests, the classification did not enter explicitly. The first set of tests is designed to predict the offspring's EEG classes from the parent's EEG classes and with contingency tables. Additionally, the correlation between the personal channel patterns (PCPs; for definition, see next section) of pairs of siblings was taken as a basis for structural analysis using a simple model computed with LISREL along the lines suggested by Heath et al. [1989]. This particular model includes only three factors to describe the outcome of a pair: a genetic factor g (common to both individuals), a stochastic factor s, and an environmental factor e (common to both individuals again). These three factors allow for the estimation of h through

$$h^2 = \frac{g^2}{(g^2 + e^2 + s^2)}$$

Normally, h is referred to as the heritability coefficient [cf. Falconer, 1961].

Tests without using any classification checked (1) for significant correlations between the spectral parameters of siblings and (2) for significant similarities [Stassen et al., 1988]. For both tests, (1) and (2), pairs had been excluded when differences on REL_B3 (relative power of frequency band three, cf. Table I for definition) of a pair did not belong to the inner two quarters of all pair differences (sequentially ordered) of the sample (V. Dellea, personal communication). Including all channels used for analysis, 18 pairs remained after this step.

Mendelian calculations are based on the assumption of equiprobability between alleles, i.e., an allele frequency of 0.5. Conditional probabilities for the children's EEG class given those of the parents can be estimated through assigning the parents the possible allele combinations. Then the occurrence of EEG classes within the offspring is found using standard combinatorics, and requiring knowledge of both parent's EEG class. The sum of possible outcomes found this way gives the conditional expectation values for the EEG classes of the children. This procedure requires knowledge of both parent's EEG class which is fulfilled for eight families.

OUTLINE OF CLASSIFICATION ALGORITHMS

Any useful classification algorithm, e.g., discriminant analysis or cluster analysis, must be independent of the local properties of a data set. One method to achieve this independence with a given data set is the method of successive random splittings [Scheidegger, 1992]. The essential idea here is to split up a basic sample of feature vectors of dimension m and size N into two disjoint subsamples of size N/2 each. This is repeated several times at random. One then tries to find patterns (in cluster analysis) or feature vectors (in discriminant analysis) which, on average, allow the same classification for all subsamples. Hence this algorithm does not find the best configuration for the N sample but rather the most reproducible one. To generalize a result based on classification, the latter property is more important than fitting an optimum to a given configuration. This is, of course, a well known problem [Krueger et al., 1981].

There are some technical differences between the applications in discriminant analysis and cluster analysis. The first follows the so called "average trajectory algorithm" (ATA) [Schmid et al., 1991; Schmid, 1993]. Basically, the ATA consists of 6 steps: (1) supervised decomposition into outcome groups (e.g., clinical EEG classification); (2) random splitting into two disjoint subsamples, the so-called learn and test samples; (3) discriminant analysis on the learn sample and application of the discriminant coefficients onto the test sample; (4) repetition of steps (2) and (3) 12 times; (5) average trajectory loop: calculation of average test and learning performances over all 12 repetitions; and (6) if the average test performance does not satisfactorily approximate the average learn performance, remove unsalient variables, i.e., variables with low averaged discriminance coefficients from the feature vector and

TABLE I. Means and Standard Deviations of Scalar EEG Parameters of the Four Clinical EEG Groups*

	Low voltage EEG		Borderline α EEG		Diffuse β EEG		Monomorphous α EEG	
Variable	μ	σ	μ	σ	μ	σ	μ	σ
Power	9.51	1.78	11.02	3.69	12.17	2.32	19.93	2.31
Abs_B1	16.81	5.18	18.23	6.32	16.06	3.64	23.34	7.08
Rel_B1	21.89	3.62	20.77	4.39	16.67	3.17	14.87	3.67
Cen_B1	2.28	0.13	2.29	0.13	2.32	0.09	2.32	0.12
Sym_B1	31.56	11.14	32.64	12.43	34.48	9.10	35.38	11.33
Freq_B1	2.62	0.64	2.60	0.62	2.64	0.58	2.71	0.62
Amp_B1	31.63	11.14	33.92	14.16	29.39	6.81	44.31	15.78
Abs_B2	15.31	3.42	17.32	6.83	16.96	3.78	28.82	8.85
Rel_B2	20.12	2.39	19.61	3.45	17.55	2.93	18.21	3.60
Cen_B2	5.51	0.09	5.53	0.11	5.57	0.10	5.70	0.15
Sym_B2	−7.78	7.71	−6.42	8.75	−3.83	8.15	5.66	11.36
Freq_B2	4.66	1.04	4.84	1.21	5.17	1.23	6.10	1.38
Amp_B2	23.58	6.10	27.06	11.93	25.36	6.35	46.98	21.42
Abs_B3	10.26	2.09	13.83	5.16	14.34	3.82	37.71	12.59
Rel_B3	26.20	2.93	30.32	5.03	28.56	4.64	45.28	5.91
Cen_B3	10.92	0.17	10.83	0.28	10.99	0.24	10.22	0.33
Sym_B3	−13.19	7.62	−19.96	13.66	−12.35	12.00	−51.86	10.91
Freq_B3	9.04	1.55	9.81	1.59	9.81	1.73	9.55	0.86
Amp_B3	17.54	4.80	29.16	15.18	26.90	13.49	136.37	65.68
Abs_B4	6.27	1.30	6.76	2.68	9.28	2.49	9.21	2.87
Rel_B4	35.13	4.32	32.60	3.97	40.25	5.65	24.77	4.06
Cen_B4	22.40	0.40	22.15	0.53	22.31	0.51	21.52	0.47
Sym_B4	−16.94	7.84	−21.55	9.78	−19.15	9.62	−34.62	8.97
Freq_B4	18.06	3.04	17.87	2.94	18.29	2.92	18.04	2.03
Amp_B4	11.58	2.66	13.47	6.24	18.31	5.45	21.65	7.58

*Parameters are labelled in the following manner: X = contribution from frequency band no. X, (X = 1..4); ABS_BX = absolute power; REL_BX = relative power; CEN_BX = centroid; SYM_BX = symmetry; AMP_BX = peak amplitude; FRQ_BX = peak frequency; VA(X) = relative peak amplitude; VF(X) = relative peak frequency.

repeat steps (2) to (5) until no further improvement can be found.

The cluster analysis [Scheidegger, 1992] consists essentially of: (1) randomly splitting the basic sample into eight pairwise disjoint subsamples of size N/2; (2) constructing distance matrices; (3) reducing the dimension through multidimensional scaling [Kruskal, 1977; Stassen et al., 1983; Angst et al., 1983]; and (4) finding prototypes with the help of the SAS procedure "Fastclus" under the supervised choice of proper seeds and cluster ordering. Step (4) is repeated until eight approximately equal solutions are found. In a next step, the original overall sample is clustered until the solution closest to the eight-subsample solution is found. The optimal feature vector starting from the original m variables is then found using the ATA. Except for step (4) of the cluster analysis, all steps of the discriminant and cluster analyses require routines from the MASTER.EEG package.

Clustering was carried out on feature vectors of a single channel. For the classification of persons, an additional step has been introduced. Accordingly we constructed a new feature vector from the classification of the channels. The relative occurrence of a certain EEG-class Y among all channels of a person was taken as the value of coordinate Y of the person's new feature vector. For example, for a person having 20% of all channels classified as 1, 20% as 2, 60% as 3 and 0% as 4, the feature vector (20,20,60,0) is assigned. With these new feature vectors, we started the clustering procedure again, searching for four types of patterns which classify persons. To avoid confusion, we will refer to this new type of feature vector as "personal channel pattern," PCP. The empirical classification found this way is also referred to as PCP-based classification throughout the remainder. In case of more than five available channels, a sliding window technique (window length: 5 channels) was applied which enabled construction of several PCPs from the same person, thus allowing a multiple classification of each person. Demanding that a person be uniquely classified into the same group via the subsequent cluster analysis provides an additional test on the robustness of our approach. In general, the greatest occurrence of a class in the feature vector also determined the classification of a person (i.e., in the example above, the person would be classified both ways into group 3). This is a confirmation of the idea that most information about a person's EEG should already be available in any given channel [Stassen, unpublished].

RESULTS

Classification Schemes

We carried out analyses of means and variances of the frequency parameters on the clinical classification scheme. Results showed differences between the group means, but also indicated overlapping of the groups (cf. Table I). Thus, except for the monomorphous α-EEG, one would not expect very good discrimination between the different clinical groups. Indeed, only those persons classified as carriers of monomorphous α EEG could be clearly separated from the rest (performance 99%), whereas the other groups showed rather marignal sep-

aration (performances: 60–75%) from one another. The optimal feature vector to discriminate clinical groups consisted of 13 variables (cf. Table IIa). This bad performance has been confirmed when we tried to discriminate on the basis of similarities [Stassen, 1985].

Because of the good distinction between the monomorphous α EEG group and the rest, we took this group to also be one of the four new groups and searched for a classification into these four groups. The new empirical classification allowed a much better discrimination (performances 87–91%) for the index cases. An eight-dimensional feature-vector was found to be optimal in this case (Table IIa). These feature vectors have been used as a basis for building PCPs to classify index cases as well as relatives.

On the average, we found about 5.5 channels per person to be classified into each person's EEG group. This means that, on the average, 80% of a person's useable channels have the same classification as that person himself. Hence, a quick, but still very good classification of a person can be done choosing the class of the maximum occurrence in the person's PCP. These findings are in agreement with our statements above.

When comparing the two classification schemes, we found the correlation between clinical and empirical classes 1 to 3 to be significant, but only 0.27 (Cramér index), while, due to construction, the monomorphous α class remains identical (cf. Table IIb). There is still good agreement between the empirical class 1 and the clinical β EEG, but the rest tends to be rather diffuse. From Table IIb, one gets the impression that the low voltage EEG splits up into two subgroups, the one sharing some properties with the clinical β group, the other and larger sharing some properties with a subgroup of the clinical borderline α group. As a main result, our analysis shows the existence of a reliable EEG classification procedure in the frequency domain but at the expense of losing information in the clinical sense.

Evidence for a Genetic Mechanism

To investigate a genetic mechanism, we've taken the first degree relatives of the clinical low voltage index cases into account. In particular, we found the following results from the parent-offspring relations (cf. Table III) within the PCP classification scheme: (a) the EEG group of the mother is a good predictor of the EEG groups of her daughters (88%) and a moderate predictor for those of her sons (40%). The EEG group of the father seem to play no role; (b) only two types of EEG groups occur within the siblings. This leads to the hypothesis of a strong mother-offspring relationship and has been confirmed by a contingency analysis of families (without fathers) vs. EEG groups: Cramér Index: 0.61, significant at the 95% level according to the Haldane-Dawson small sample test [Bortz et al., 1990].

Structural analysis with LISREL lead to heritability coefficients h in the range 0.65–0.73 depending on the genetic correlation assumed in the model. The latter is normally set to 0.5, but can be higher in the case of genetic dominance. Increasing the genetic correlation in the model has led to higher values of h with our sample.

The classification-independent measures showed a picture consistent with the findings above. We found significant nonzero correlations (95% level) between pairs of siblings for up to 12 frequency parameters. A high proportion of them (40–45%) stems from the α-band (band no. 3). In the same way, similarities have been shifted from $(\mu_S, \sigma_S) = (0.73, 0.12)$ for the sample population (corrected for multiple occurrences of pairs) to (0.82, 0.07) for siblings (cf. Fig. 2).

Preferences

On the one hand, the results of our approach show evidence for a genetic mechanism for the low voltage EEG. On the other hand, we have found differences when including the fathers of the index cases into the genetic analysis or when increasing genetic correlation within structural analysis. Hence, the question arises as to what extent one has to deal with differences between males and females within this context. The most simple differences can be expected when males and females show different preferences within the subgroups. Therefore, we searched for differences within the ratios of PCP-based Type 1 and Type 3 EEG as compared by sex. Two-by-two contingency analysis, including Yates small sample correction [Yates, 1934], showed significant differences (95% level) for both males and females. Males tend to prefer Type 3 EEG and females Type 1 EEG. This can be expected when looking at Table III. However, when the fathers are included into the con-

TABLE IIa. Relation Between Empirical and Clinical Classification of Index Cases as Found With the ATA* Performances From Discriminant Analysis for Clinical and Empirical EEG Groups 1 to 3 and Salient Variables Building the Final Feature Vector

	Performance of group no.			
Classification	1	2	3	Salient variables
Old (clinical)	0.70	0.60	0.75	ABS_B1 REL_B1 ABS_B2 REL_B2[a] ABS_B3 ABS_B4 REL_B4 SYM_B4 VA(2) VF(2) VA(3) VA(4) Power
New (PCP-based)	0.89	0.91	0.87	ABS_B1 REL_B1 ABS_B2 ABS_B3 REL_B3 ABS_B4 REL_B4 VA(2)

*Group 4 (monomorphous α EEG) is identical and disjoint to the rest by construction.
[a]See Table I for meaning of the labels.

TABLE IIb. Relation Between Old (Clinical) and New (Empirical) Groups*

Old group	1	2	3	4
	New group			
1	4	3	8	0
2	0	2	1	0
3	11	13	3	0
4	0	0	0	18

*Groups *1 to 3*: $\chi^2 = 13.9$, df = 4, $P < 0.01$, Correlation (Cramér Index): 0.27

tingency tables, the significance drops to the 90% level. We will discuss this in more detail below.

A Mendelian Two-Gene Interpretation

Our sample consists of four classes which occur with a 7:3:17:2 ratio (PCP-based classification) within the offspring. This is not significantly different ($\chi^2 = 1.70$, df = 3) from a 3:3:9:1 ratio. Hence the simplest Mendelian scheme which is statistically compatible with these facts includes two genes A,a; B,b. The capitals stand for the dominant and the lowercase letters for the recessive occurrence of the gene. To test this hypothesis further, we assigned the conditional probabilities for the children's EEG class, given those of the parents. Within our data, the best fit in terms of such a simple Mendelian model has been found with the following phenotype-genotype assignment: 1: AAbb, Aabb; 2: aaBB, aaBb; 3: AABB, AABb, AaBB, AaBb, 4: aabb; where we identify the phenotype with the empirical EEG class. This working hypothesis leads to a χ^2 of 4.62 (Table IV) between model and data (insignificant for df = 3) and a correlation of $\rho = 0.85$ between model and EEG types. This is obtained only from the data of the eight families from which both parents have been available (see above). Hence our hypothesis of a simple genetic mechanism cannot be rejected.

For the sake of completeness, we mention that the major contribution to χ^2 (ca. two-thirds) stem from the positive deviation of the Type 3 EEG from the model.

TABLE III. Empirical Classification of Clinical Low Voltage EEG Index Cases and Their Relatives Up to First Order

	Parents		Offspring	
Family	Male	Female	Male	Female
1	1	2	3	2
3	1	2	1	2
4	n.o.	1	3	
5	n.o.	2	3	
6	n.o.	1	3 3	1 1 1
7	2	1	1	5 1
9	1	3	3	
10	1	2	3 2	
11	n.o.	1	1 3 3 3	
12	5	1	3	
13	5	3	3 3	
15	2	3	3 3 3 5	3

DISCUSSION

Our results show strong agreement between a person and him- or herself. This is clearly demonstrated by the fact that an (hypothetical) "average" person has 6.6 channels, from which 5.5 would belong to the same EEG class. (A natural person cannot have a noninteger number of available channels, of course.) This finding holds, even though no information about the person from which a particular channel stems enters into the channel classification scheme. Hence, most of the information about a person's EEG seems to be contained within the information obtainable from a single channel. This is in good agreement with other findings [Stassen, unpublished].

With regard to discriminant performance, our findings seem at first glance to contradict claims of easy discrimination between clinical EEG classes [Vogel, 1986]. However, it has been shown in Table I that the clinical groups do indeed overlap, in agreement with earlier findings [Propping et al., 1980]. Nevertheless our findings do cast some doubt about the existence of an easy and reliable discrimination for clinical classification.

There are two differences inherent to our approach as compared to other approaches: firstly, we discriminate within the frequency domain and not within the time domain. This has computational advantages, as already stated above. Secondly, our approach searches for as global a solution as possible. By this, we mean the following: given a sample of size N representative for the total population, our algorithm tries to optimize according to the properties of the total population and not to the properties of a specific N sample. Indeed, failures of traditional discriminant procedures in the time domain have been found, when applied to different samples of EEG probands [Krueger et al., 1981].

The specific influence of changing to the time domain could not be studied. Therefore, we do not know exactly to what extent the weak correlation between clinical and empirical classes originates from this technical detail. With respect to clinical groups, however, our results plus the above cited results suggest the global nonreliability rather than the strength of any automatic discrimination procedure regardless of whether the analysis is carried out in the time domain or in the frequency domain. It is in view of this fact that one must, however, investigate properties of EEG classes found with different classification schemes. Since we could not explore this problem in full detail, we focused our attention on the genetic problem of the low voltage EEG. We thereby did not search for different explanation for the underlying genetic mechanism but we simply tried to demonstrate the consistency of our scheme with the Mendelian genetic model proposed by Vogel [1986].

Strong evidence for a genetic mechanism within low voltage EEG has indeed been found in terms of significant correlations within families. Furthermore, the observed increase of the mean and the reduction of the standard deviation in the similarity distribution of the siblings as compared to the whole sample is typical for

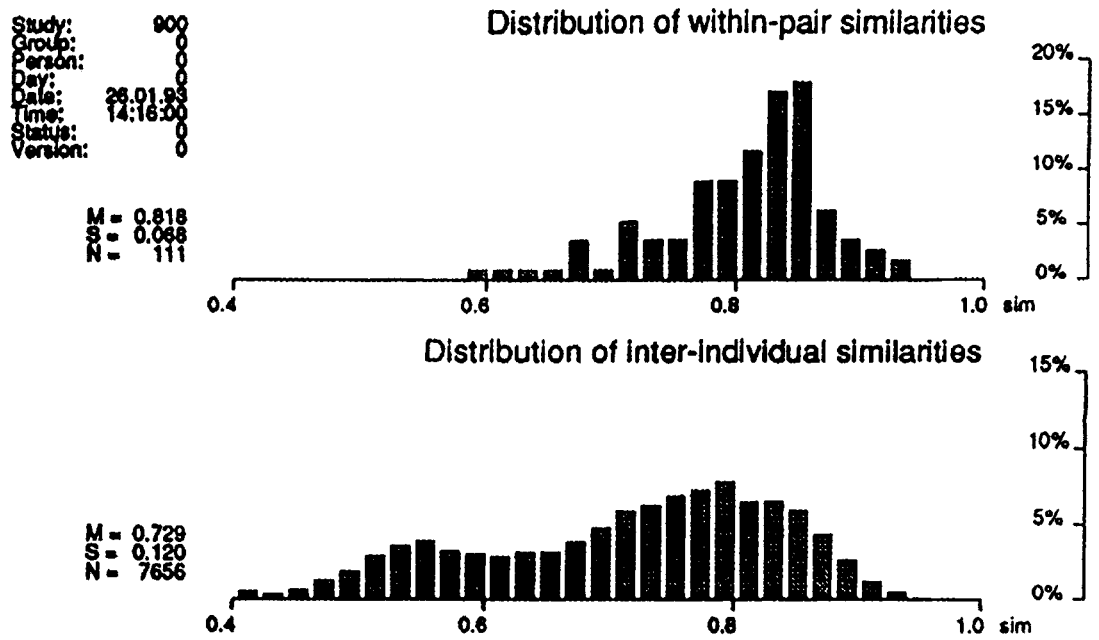

Fig. 2. Distribution of within-pair similarities of siblings **(upper)** and distribution of interindividual similarities of total sample **(lower)**.

persons genetically closely related [Stassen, 1991]. Our higher values found for the heritability coefficient h are in excellent agreement with earlier findings on human EEG [Stassen et al., 1993]. As stated above, these higher values occurred when increasing the genetic correlation in the model. This can happen when genetic dominance needs to be considered [Heath et al., 1989]. This possibility of the phenotype level might be a hint to the existence of a genetic process including dominant alleles on the genotype level.

The search for evidence of a genetic mechanism also led us in a natural way to study differences between males and females. In particular, the phenotype of the fathers did not fit well into our schemes. This striking fact needs some discussion. The age of the fathers might play a possible role here. It has been claimed that older persons whose EEG belongs to the low voltage group tend to share more properties with β EEG. The same seems to be true for females as compared to males [Anokhin et al., 1992]. Hence, one cannot exclude the possibility that the high proportion of fathers belonging to the empirical (hence PCP-based) Class 1 actually belonged to the empirical Class 3 some decades ago. Due to the tendency for males to belong to the empirical Class 3, the fathers would fit reasonably well within our scheme in that case, except for the failure to predict the daughter's EEG class when we know the father's EEG class. Furthermore, one can only speculate on the classification of the unknown fathers. But the observed preferences are consistent with our interpretations (a) that empirical Type 1 EEG share some basic properties with clinically diffuse β EEG, and (b) that persons with empirical Type 3 EEG have basic properties in common with those belonging clinically to a subgroup of the borderline α EEG.

On the one hand, aging effects on the EEG are discussed somewhat controversially [Stassen and Bomben, unpublished, and refs. therein]. On the other hand, the basic genetic mechanism has been recently given a more profound basis [Steinlein et al., 1992]. Bearing in mind these facts, we also searched for another explanation. So far, we treated the genes A,a and B,b in the traditional Mendelian sense. A shift towards more consistency when including fathers would be to interpret BB,Bb,bb in a rather polygenic sense, where a threshold somewhere between Bb and bb would be assumed, allowing for a better fit of theoretical occurrences to observed ones, especially to empirical Types 1 and 3. The latter showed the highest contribution to χ^2, as stated above. This view would be along the lines suggested by Vogel [1986]: absence of one dominant gene (A) gives rise to empirical Type 2 or Type 4 EEG, which are nearly disjoint to the clinical low voltage group (Table II). On the other hand, empirical Type 1 (correlated with the clinical β EEG) and empirical Type 3 would be determined by a polygenic mechanism. Grouping empirical Type 2 and empirical Type 4 together seems not unreasonable because, in two mothers the PCP con-

TABLE IV. Expected Frequencies According to the Mendelian Model and Observed Frequencies of Offspring*

Frequencies	Group 1	Group 2	Group 3	Group 4
Expected	5	4	6	3
Observed	3	3	10	2

*Only those 8 families have been included from which the EEG group of both parents are known (see text). Nearest integers to expected occurrences are shown. Equiprobability has been assumed for the distribution of the genes A,a,B,b. χ^2: 4.62, df = 3, $P > 0.05$; correlation between the two lines: $\rho = 0.85$.

tained Type 2, when classified empirically, but also some Type 4 when discriminated with the clinical feature vector. We remember that clinical Type 4 EEG (α EEG), as derived from the index cases, has been set identically to empirical Type 4. We mention in passing that (because of the data set size) we cannot completely exclude nonpolygenic mechanisms, e.g., a sex-related mechanism where, e.g., maternal effects are considered (M. Ribi, personal communication), or in the case that the recessive genes might be unequally distributed between males and females. These latter views, however, do hardly lead to compatibility with the more recent results cited above. As a basic fact, we suggest interpreting the genetic mechanism of the low voltage EEG along the proposed lines, although empirically we get a classification scheme which differs from the clinical classification. This confirms our view that a basic mechanism related to EEG classes must be independent of the classification scheme actually used for its description.

CONCLUSION

We found an algorithm for a reliable classification of EEG groups. Mainly (but not only) based on frequency tables, we have been able to reproduce a basic genetic mechanism for the low voltage EEG irrespective of the classification scheme actually used, although our data did not allow definitive answers to questions arising on sex or age effects. However, the fact that the proposed genetic mechanism does not seem to be strongly dependent on the particular classification scheme actually used confirms earlier findings. This lends confidence to results obtained with other classification schemes though they are not reliable in the sense defined in the Introduction.

ACKNOWLEDGMENTS

We thank Dr. M. Ribi, Zoological Institute, Dr. E. Pavesi and V. Dellea, Psychiatric University Hospital, all from Zürich University, and Dr. C. Stärk from the Cantonal Psychiatric Hospital "Insel," CH-8462 Rheinau, for their important remarks during the preparation of this manuscript.

REFERENCES

Angst J, Scharfetter C, Stassen HH (1983): Classification of Schizoaffective patients by multidimensional scaling and cluster analysis. Psychiatria Clin 16:254–264.

Anokhin A, Steinlein O, Fischer C, Yping M, Vogt P, Schalt E, Vogel F (1992): A genetic study of the human low voltage electroencephalogram. Hum Genet 90:99–112.

Bortz J, Lienert GA, Boehnke K (1990): "Verteilungsfreie Methoden in der Biostatistik." Berlin: Springer Verlag.

Cooper R, Osselton JW, Shaw JC (1969): "EEG Technology." London: Butterworths.

Falconer DS (1961): "Introduction to Quantitative Genetics." Edinborough/London: Oliver and Boyd.

Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fulker DW (1989): Testing structural equation models for twin data using LISREL. Behav Genet 19:9–35.

Jöreskog KG, Sörbom D (1989): "LISREL User's Reference Guide." Mooresville: Scientific Software Inc.

Krüger J, Schalt E, Vogel F (1981): Charakterisierung erblicher EEG-varianten mit Hilfe der Amplituden-Intervallanalyse. Z EEG-EMG 12:113–119.

Kruskal JB (1977): Multidimensional scaling and other methods for discovering structure. In Enslein K, Ralston A, Wilf HS (eds): "Statistical Methods for Digital Computers." New York: Wiley, pp. 296–339.

Propping P, Krüger J, Janah A (1980): Effect of alcohol on genetically determined variants of the normal electroencephalogram. Psychiatr Res 2:85–89.

Scheidegger P (1992): "Reproduzierbarkeit von Strukturen in empirischen Daten." PhD Thesis Univ. Zürich, Zentralstelle der Studentenschaft Zürich.

Schmid GB (1993): The Average Trajectory Algorithm. Appendix I in (Dünki, 1993, "Zur Vererbung des Niederspannungs-EEG." Internal Research Report Psych. Univ. Hosp. Zürich).

Schmid GB, Stassen HH, Gross G, Huber G, Angst J (1991): Long term prognosis of schizophrenia. Psychopathology 24:130–140.

Stassen HH (1995): The similarity approach to EEG analysis. Meth Inform Med 24:200–212.

Stassen HH (1991): The octave approach to EEG analysis. Meth Inform Med 30:104–110.

Stassen HH (1992): "Master.EEG Programs for Investigations into the Properties of EEG Time Series." User Manual on Version 7.1, Psychiatric University Hospital Zürich.

Stassen HH: "How Individual are Human Brain Wave Patterns?" (Monograph in prep.).

Stassen HH, Bomben G: Longterm stability of human brain wave pattern. (In prep.).

Stassen HH, Günter R, Bomben G (1983): Multidimensionale Skalierung: Gewinnung metrischer Information aus nicht-metrischen Daten am Beispiel von EEG Spektralmustern. In Kazmiejczak H (ed): "Mustererkennung 1983. Vortrge des 5. DAGM-Symposiums, Karlsruhe 1983." VDE-Fachberichte 35, VDE Verlag Berlin-Offenbach, pp. 107–111.

Stassen HH, Bomben G, Propping P (1987): Genetic aspects of the EEG: An investigation into the within-pair similarity of monozygotic and dizygotic twins with a new method of analysis. Electroencephal Clin Neurophys 66:489–501.

Stassen HH, Lykken DT, Propping P, Bomben G (1988): Genetic determination of the human EEG. Hum Gen 80:165–176.

Stassen HH, Lykken DT, Propping P (1993): Zwillingsuntersuchungen zur Genetik des normalen Elektroenzephalogramms. In Baumann P (ed): "Biologische Psychiatrie der Gegenwart." Wien: Springer-Verlag, pp 139–144.

Steinlein O, Anokhin A, Yping M, Schalt E, Vogel F (1992): Localization of a gene for the human low-voltage on 20q and genetic heterogeneity. Genetics 12:69–73.

Vogel F (1986): Grundlage und Bedeutung genetisch bedingter Variabilität des normalen menschlichen EEG. Z EEG-EMG 17:173–188.

Wright JJ (1990): Reticular activation and the dynamics of neural networks. Biol Cybern 62:289–298.

Yates F (1934): Contingency tables involving small numbers and the χ^2-test. J Royal Statist Soc Suppl 1:217–235.